# Coin Flipping with Constant Bias Implies One-Way Functions

Iftach Haitner[*] Eran Omri[†]

October 20, 2011

**Abstract**

It is well known (cf., Impagliazzo and Luby [FOCS '89]) that the existence of almost all "interesting" cryptographic applications, i.e., ones that cannot hold information theoretically, implies one-way functions. An important exception where the above implication is not known, however, is the case of coin-flipping protocols. Such protocols allow honest parties to mutually flip an unbiased coin, while guaranteeing that even a cheating (efficient) party cannot bias the output of the protocol by much. Impagliazzo and Luby proved that coin-flipping protocols that are safe against negligible bias do imply one-way functions, and, very recently, Maji, Prabhakaran, and Sahai [FOCS '10] proved the same for constant-round protocols (with any non-trivial bias). For the general case, however, no such implication was known.

We make progress towards answering the above fundamental question, showing that (strong) coin-flipping protocols safe against a constant bias (concretely, $\frac{\sqrt{2}-1}{2} - o(1)$) imply one-way functions.

**Keywords:** coin-flipping protocols; one-way functions;

# 1 Introduction

A central focus of modern cryptography has been to investigate the weakest possible assumptions under which various cryptographic primitives exist. This direction of research has been quite fruitful, and minimal assumptions are known for a wide variety of primitives. In particular, it has been shown that one-way functions (i.e., easy to compute but hard to invert functions) imply pseudorandom generators, pseudorandom functions,

[*] School of Computer Science, Tel Aviv University. E-mail: iftachh@cs.tau.ac.il. Research partly supported by the Check Point Institute for Information Security

[†] Department of Computer Science, Bar Ilan University. E-mail: omrier@gmail.com. Research supported by the European Research Council as part of the ERC project LAST.

symmetric-key encryption/message authentication, commitment schemes, and digital signatures [10, 11, 14, 13, 21, 22, 23], where one-way functions were shown also to be implied by each of these primitives [16].

An important exception for which we have failed to prove the above rule, is that of coin-flipping protocols. A coin-flipping protocol [3] allows the honest parties to mutually flip an unbiased coin, where even a cheating (efficient) party cannot bias the outcome of the protocol by much. While one-way functions are known to imply coin-flipping protocols [3, 9], the other direction is less clear: Impagliazzo and Luby [16] showed that *negligible*-bias coin-flipping protocols (i.e., an efficient cheating strategy cannot make the common output to be 1, or to be 0, with probability greater than $\frac{1}{2}+\operatorname{neg}(n)$) implies one-way functions. Very recently, Maji, Prabhakaran, and Sahai [18] proved the same implication for $(\frac{1}{2}-1/\operatorname{poly}(n))$-bias *constant-round* protocols, where $n$ is the security parameter of the protocol. We have no such implications, however, for any other choice of parameters.

## 1.1 Our Result

We prove the following theorem.

**Theorem 1** (informal)**.** *The existence of a* $(\frac{\sqrt{2}-1}{2}-o(1))$*-bias coin-flipping protocol (of any round complexity) implies one-way functions.*[1]

## 1.2 Related Results

As mentioned above, Impagliazzo and Luby [16] showed that negligible-bias coin-flipping protocols imply one-way functions, and Maji et al. [18] proved the same for $(\frac{1}{2}-1/\operatorname{poly}(n))$-bias constant-round protocols. [18] also proved that of $(\frac{1}{4}-o(1))$-bias coin-flipping protocols implies that BPP $\neq$ NP. Finally, it is well known that $(\frac{1}{2}-\upsilon(n))$-bias coin-flipping protocols, for any $\upsilon(n)>0$, implies that BPP $\neq$ PSPACE. All the above results extend to *weak* coin-flipping protocols: in such protocols, each party has a *different* predetermined value towards which it cannot bias the output coin.[2] A quick overview on the techniques underlying the above results, can be found in Section 1.3.3.

*Information theoretic* coin-flipping protocols (i.e., whose security holds against all powerful adversaries) were shown to exist in the quantum world; Mochon [19] presents an $\varepsilon$-bias quantum weak coin-flipping protocol for any $\varepsilon>0$. Chailloux and Kerenidis [4] present a $\left(\frac{\sqrt{2}-1}{2}-\varepsilon\right)$-bias quantum strong coin-flipping protocol for any $\varepsilon>0$ (which is optimal, [17]). A key step in [4] is a reduction from strong to weak coin-flipping protocols, which holds also in the classical world (see Section 6 for further discussion).

A related line of work considers *fair* coin-flipping protocols. In this setting the honest party is required to always output a bit, whatever the other party does. In particular, a

[1] We note that our results do not apply to weak coin-flipping protocols. See Section 6 for further discussion.

[2] While such protocols are strictly weaker then full-fledged coin flipping protocols, they are still useful in many settings. For instance, when Alice and Bob are trying to decide who is doing the dishes.

cheating party might bias the output coin just by aborting. We know that one-way functions imply fair $(1/\sqrt{m})$-bias coin-flipping protocol [1, 6], where $m$ being the round complexity of the protocol, and this quantity is known to be tight for $O(n/\log n)$-round protocols with fully black-box reductions [8]. Oblivious transfer, on the other hand, implies fair $1/m$-bias protocols [20, 2], which is known to be tight [6].

## 1.3 Our Technique

Let $(\mathsf{A},\mathsf{B})$ be a balanced coin-flipping protocol (i.e., the common output of the honest parties is a uniformly chosen bit), and let $f$ be the following efficiently computable function:

$$f(r_{\mathsf{A}}, r_{\mathsf{B}}, i) = \mathrm{Trans}(r_{\mathsf{A}}, r_{\mathsf{B}})_i, \mathrm{Out}(r_{\mathsf{A}}, r_{\mathsf{B}})$$

where $r_{\mathsf{A}}$ and $r_{\mathsf{B}}$ are the random coins of $\mathsf{A}$ and $\mathsf{B}$ respectively, $\mathrm{Trans}(r_{\mathsf{A}}, r_{\mathsf{B}})_i$ is the first $i$ messages exchanged in the execution $(\mathsf{A}(r_{\mathsf{A}}), \mathsf{B}(r_{\mathsf{B}}))$, and $\mathrm{Out}(r_{\mathsf{A}}, r_{\mathsf{B}})$ is the common output of this execution (i.e., the coin). Assuming that one-way functions do not exist, it follows that distributional one-way functions do not exist either [16], and therefore there exists an efficient inverter $\mathsf{Inv}$ that given a random output $y$ of $f$, samples a random preimage of $y$. Concretely, for any $p \in \mathrm{poly}$ there exists a PPT $\mathsf{Inv}$ such that the following holds:

$$\mathrm{SD}((X, f(X)), (\mathsf{Inv}(f(X')), X')) \leq 1/p(|X|) \tag{1}$$

where $X$ and $X'$ are uniformly distributed over the domain of $f$, and SD stands for statistical distance. In the following we show how to use the above $\mathsf{Inv}$ to bias the output of $(\mathsf{A},\mathsf{B})$.

Note that given a random partial transcript $t$ of $(\mathsf{A},\mathsf{B})$, the call $\mathsf{Inv}(t,1)$ returns a random pair of random coins for the parties that is (1) consistent with $t$, and (2) yields a common output 1. In other words, one can use $\mathsf{Inv}$ to sample a random continuation of $t$ which leads to a 1-leaf — a full transcript of $(\mathsf{A},\mathsf{B})$ in which the common output is 1. As we show below, such capability is an extremely useful tool for a dishonest party trying to bias the outcome of this protocol. In particular, we consider the following cheating strategy $\mathcal{A}$ for $\mathsf{A}$ (a cheating strategy $\mathcal{B}$ for $\mathsf{B}$ is analogously defined): given that the partial transcript is $t$, $\mathcal{A}$ uses $\mathsf{Inv}$ to sample a pair of random coins $(r_{\mathsf{A}}, r_{\mathsf{B}})$ that is consistent with $t$ and leads to a 1-leaf ($\mathcal{A}$ aborts if $\mathsf{Inv}$ fails to provide such coins), and then acts as the honest $\mathsf{A}$ does on the random coins $r_{\mathsf{A}}$, given the transcript $t$. Namely, at each of its turns $\mathcal{A}$ takes the first step of a random continuation that leads to a 1-leaf.

Assuming that $\mathsf{Inv}$ behaves as its ideal variant that returns a uniform random preimage on *any* transcript, it is not that hard to prove (see outline in Section 1.3.1) that either $\mathcal{A}$ or $\mathcal{B}$ can significantly bias the outcome of the protocol. Proving that the same holds with respect to the real inverter, however, is not trivial. Algorithm $\mathsf{Inv}$ is only guaranteed to work well on *random* transcript/output pairs, as induced by a random output of $f$ (namely, a transcript/output pair defined by a random *honest* execution of $(\mathsf{A},\mathsf{B})$). A random execution of $(\mathcal{A},\mathsf{B})$ or of $(\mathsf{A},\mathcal{B})$ (i.e., with one party being controlled by the adversary) might, however, generate a query distribution that is very far from that induced by $f$.

Fortunately, we manage to prove (and this is the crux of our proof, see outline in Section 1.3.2) that the following holds: We call a query *non-typical*, if its probability mass with respect to the execution of $(\mathcal{A}, \mathsf{B})$ (or of $(\mathsf{A}, \mathcal{B})$) is much larger than its mass with respect to the output distribution of $f$. We first show that even if both $\mathcal{A}$ and $\mathcal{B}$ totally fail on such non-typical queries, then either $\mathcal{A}$ or $\mathcal{B}$ can significantly bias the outcome of the protocol assuming access to the ideal sampler. Since on typical queries the real sampler should perform almost as well as its ideal version, we conclude that the cheating probability of either $\mathcal{A}$ or $\mathcal{B}$ is high, also when the cheating strategies are using the real sampler.

### 1.3.1 When Using the Ideal Sampler

Consider a mental experiment in which the cheating strategies $\mathcal{A}$ and $\mathcal{B}$ (both using the ideal sampler) are interacting with each other. It is not hard to see that the common output of $(\mathcal{A}, \mathcal{B})$ in this case is always one. Moreover, the transcript distribution induced by such an execution, is that of a random execution of the "honest" protocol $(\mathsf{A}, \mathsf{B})$ conditioned that the common output is 1 (i.e., a random 1-leaf). In particular, the probability of each 1-leaf in a random execution of $(\mathcal{A}, \mathcal{B})$ is twice its probability in $(\mathsf{A}, \mathsf{B})$.

The probability of a 1-leaf $t$ to happen, is the product of the probabilities that in each stage of the protocol the relevant party sends the "right" message. Such a product can be partitioned into two parts: the part corresponding to the actions of $\mathcal{A}$, and the part corresponding to the actions of $\mathcal{B}$. In particular, either $\mathcal{A}$ or $\mathcal{B}$ contributes a factor of value at least $\sqrt{2}$ to the probability of $t$. Namely, the probability of a 1-leaf $t$ in either $(\mathcal{A}, \mathsf{B})$ or in $(\mathsf{A}, \mathcal{B})$, is $\sqrt{2}$ times its probability in $(\mathsf{A}, \mathsf{B})$. Summing over all 1-leaves, it follows that the common output of either $(\mathcal{A}, \mathsf{B})$ or $(\mathsf{A}, \mathcal{B})$ is one with probability at least $\sqrt{2} \cdot \frac{1}{2} = 1/\sqrt{2}$. That is, either $\mathcal{A}$ or $\mathcal{B}$ can bias the output of $(\mathsf{A}, \mathsf{B})$ by at least $\frac{1}{\sqrt{2}} - \frac{1}{2} = \frac{\sqrt{2}-1}{2}$.

### 1.3.2 Using the Real Sampler

By the discussion we made earlier, it suffices to prove that the following holds: either $\mathcal{A}$ or $\mathcal{B}$ can bias the output of the protocol significantly, when given access to the ideal sampler, even if both cheating strategies are assumed to fail completely when asking non-typical queries.

Towards this end, we partition the non-typical queries into two: (1) queries $(t, 1)$ such that the probability to visit $t$ in $(\mathcal{A}, \mathsf{B})$ or $(\mathsf{A}, \mathcal{B})$, is much larger than this probability with respect to $f$ (i.e., super polynomial in $n$ larger than $\Pr[f(X) = (t, *)]$), and (2) queries $(t, 1)$ such that the probability of ending in a 1-leaf conditioned on $t$ is small (i.e., $\Pr[f(X) = (t, 1) \mid f(X) = (t, *)]$ is small). In the following we focus on the first type of non-typical queries, which we find to be the more interesting case.

For $q \in \mathbb{N}$, let $\mathsf{UnBal}_{\mathcal{A}}$ contain the transcripts whose weights induced by $(\mathcal{A}, \mathsf{B})$ are at least $q$ times larger then their weights in the honest protocol ($\mathsf{UnBal}_{\mathcal{B}}$ is defined similarly). Using similar intuition to that used in Section 1.3.1, one can show that the probability of every transcript $t$ induced by a random execution of $(\mathcal{A}, \mathcal{B})$ is at most twice its probability in a random (honest) execution of $(\mathsf{A}, \mathsf{B})$. Hence, the following "compensation effect" happens: if the probability of a transcript $t$ in (a random execution of) $(\mathsf{A}, \mathcal{B})$ is $q$ times *larger* than its probability in $(\mathsf{A},\mathsf{B})$, then the probability of $t$ in $(\mathcal{A},\mathsf{B})$ is $q$ times *smaller* than this value. We conclude that $\mathsf{UnBal}_{\mathcal{A}}$ is visited by $\mathcal{B}^{\mathsf{Ideal}}$ with probability at most $1/q$.

To show that both $\mathcal{A}$ and $\mathcal{B}$ can be assumed to fail completely when asking queries in $\mathsf{UnBal}_{\mathcal{A}}$ (the argument for $\mathsf{UnBal}_{\mathcal{B}}$ is analogous), we consider another mental experiment. In this mental experiment, we replace the probabilities of ending up with a 1-leaf, upon reaching a transcript in $\mathsf{UnBal}_{\mathcal{A}}$ by associating a new values to each such transcript. These values are no longer probability measures. Specifically, for all $t \in \mathsf{UnBal}_{\mathcal{A}}$, we replace the probability that $(\mathcal{A},\mathsf{B})$ ends up in a 1-leaf conditioned on $t$ with the value $1/\sqrt{q}$ and replace the probability that $(\mathsf{A},\mathcal{B})$ ends up in a 1-leaf conditioned on $t$ with the value $\sqrt{q}$ (this is only a mental experiment, so we can allow these values to be larger than 1). Using a similar approach to that used in Section 1.3.1, we can prove that in the above experiment, it is still true that either $\mathcal{A}$ or $\mathcal{B}$ biases the output of $(\mathsf{A},\mathsf{B})$ by $\frac{\sqrt{2}-1}{2}$.

Finally, we note that we can safely fail both cheating strategies on $\mathsf{UnBal}_{\mathcal{A}}$ almost without changing their overall success probability in the above experiment. Specifically, $\mathcal{A}$ will not suffer much since it visits these nodes with probability at most 1 and gains only $1/\sqrt{q}$ upon visiting them. On the other hand, $\mathcal{B}$ will not suffer much since it visits these nodes with probability at most $1/q$ and gain only $\sqrt{q}$ upon visiting them (hence, these nodes contributes at most $1/\sqrt{q}$ to its overall success). Observe that the probabilities induced by an execution of $(\mathcal{A},\mathsf{B})$ (or of $(\mathsf{A},\mathcal{B})$) on typical transcripts in the real scenario, as well as, the success probability of the adversary upon visiting these transcripts, are exactly the same as in the above mental experiment. We conclude that either $\mathcal{A}$ or $\mathcal{B}$ biases the output of $(\mathsf{A},\mathsf{B})$ by $\frac{\sqrt{2}-1}{2} - 1/\operatorname{poly}$, even assuming that both cheating strategies totally fail on non-typical queries.

### 1.3.3 Perspective

The sampling strategy we use above was inspired by the "smooth sampling" approach used by [5, 12, 15] in the setting of parallel repetition of interactive arguments to sample a random wining strategy for the cheating prover. Such approach can be thought of as an "hedged greedy" strategy, using the recent terminology of Maji et al. [18], as it does not necessarily choose the *best* move at each step (the one that maximize the success probability of the honest strategy), but rather *hedges* its choice according to the relative success probability. [18] used a different hedged greedy strategy to bias any coin-flipping protocol by $\frac{1}{4} - o(1)$. They then show how to implement this strategy using an NP-oracle, yielding that $(\frac{1}{4} - o(1))$-bias coin-flipping protocols imply BPP $\neq$ NP. Their proof, however, does not follow through using a one-way functions inverter, and thus, does not yield that such protocols imply that one-way functions do not exist.

Impagliazzo and Luby [16] used a more conservative method to bias a coin-flipping protocol by $\frac{1}{\sqrt{m}}$ (where $m$ is the protocol round complexity). Their cheating strategy (which, in turn, was inspired by [7]) follows the prescribed one (i.e., acts honestly), while deviating from it at most once through the execution. In particular, at each step it estimates its potential gain from deviating from the prescribed strategy. If this gain is large enough, it deviates from the prescribed strategy, and then continues as the honest party would. Since their strategy only needs to estimates the potential gain *before* deviating from the prescribed strategy, it is rather straightforward to prove that it can be implemented using a one-way function inverter (in particular, the query distribution induced by their strategy is simply the output distribution of the one-way function).

Finally, we mention that the cheating strategy used by [18] to prove their result for constant-round protocols, takes a very different approach then the above. Specifically, their cheating strategy uses a one-way function inverter to implement (with close resemblance) the well-known recursive PSPACE-attack on such protocols. Unlike the above greedy strategies, the running time of this recursive approach is exponential in the round complexity of the protocol (which is still efficient for constant-round protocols).

## Paper Organization

General notations and definitions used throughout the paper are given in Section 2. Our adversarial strategy to bias any coin-flipping protocol is presented in Section 3. In Section 4 we analyze this strategy assuming access to an ideal sampler. Finally, in Section 5 we extend this analysis to the real sampler.

# 2 Preliminaries

## 2.1 Notation

We use calligraphic letters to denote sets, uppercase for random variables, and lowercase for values. For an integer $n \in \mathbb{N}$, we let $[n] = \{1, \cdots, n\}$.

A function $\mu \colon \mathbb{N} \to [0,1]$ is *negligible*, if $\mu(n) = n^{-\omega(1)}$, where neg denotes an arbitrary negligible function. (In particular, $f(n) = \operatorname{neg}(n)$ means that $f$ is negligible, where $f(n) > \operatorname{neg}(n)$ means that $f$ is not negligible.) We let poly denote an arbitrary polynomial, and let PPT denote the set of probabilistic algorithms (i.e., Turing machines) that run in *strict* polynomial time. Given a two-party protocol $(\mathsf{A}, \mathsf{B})$ and inputs $i_{\mathsf{A}}$ and $i_{\mathsf{B}}$, we let $\operatorname{Out}(\mathsf{A}(i_{\mathsf{A}}), \mathsf{B}(i_{\mathsf{B}}))$ and $(\mathsf{A}(i_{\mathsf{A}}), \mathsf{B}(i_{\mathsf{B}}))$ denote the (joint) output and transcript respectively, of the execution of $(\mathsf{A}, \mathsf{B})$ with inputs $i_{\mathsf{A}}$ and $i_{\mathsf{B}}$.

Given a random variable $X$, we write $x \leftarrow X$ to indicate that $x$ is selected according to $X$. Similarly given a finite set $\mathcal{S}$, we let $s \leftarrow \mathcal{S}$ denote that $s$ is selected according to the uniform distribution on $\mathcal{S}$. We adopt the convention that when the same random variable occurs several times in an expression, all occurrences refer to a single sample. For example, $\Pr[f(X) = X]$ is defined to be the probability that when $x \leftarrow X$, we have $f(x) = x$. We write $U_n$ to denote the random variable distributed uniformly over $\{0,1\}^n$. Given a measure $M$ over a set $\mathcal{S}$, the support of $M$ is defined as $\operatorname{Supp}(M) := \{s \in \mathcal{S} \colon M(s) > 0\}$. The statistical distance of two distributions $P$ and $Q$ over a finite set $\mathcal{U}$, denoted $\operatorname{SD}(P, Q)$, is defined as $\frac{1}{2} \cdot \sum_{u \in \mathcal{U}} |P(u) - Q(u)|$. We use the following notion of measure dominance.

**Definition 2** (dominating measure)**.** *A measure $M$ is said to $\delta$-*dominate *a measure $M'$, if:*

1. $\operatorname{Supp}(M') \subseteq \operatorname{Supp}(M)$, *and*

2. *$M(y) \geq \delta \cdot M'(y)$, for every $y \in \mathrm{Supp}(M')$.*

## 2.2 Coin-Flipping Protocols

In a coin-flipping protocol the honest execution outputs an unbiased coin, where no (efficient) cheating party can bias the outcome by much. This intuitive description is captured using the following definition.

**Definition 3.** *A polynomial-time protocol $(\mathsf{A},\mathsf{B})$ is a $\delta$-bias coin-flipping protocol, if the following hold:*

1. $\Pr[\mathrm{Out}(\mathsf{A},\mathsf{B})(n) = 0] = \Pr[\mathrm{Out}(\mathsf{A},\mathsf{B})(n) = 1] = \frac{1}{2}$, *and*
2. *for any* PPT*'s $\mathcal{A}$ and $\mathcal{B}$, any $c \in \{0,1\}$ and all large enough $n$:*
   $\Pr[\mathrm{Out}(\mathcal{A},\mathsf{B})(n) = c], \Pr[\mathrm{Out}(\mathsf{A},\mathcal{B})(n) = c] \leq \frac{1}{2} + \delta(n)$.

*In the case that $\delta(n) = \mathrm{neg}(n)$, we simply say that $(\mathsf{A},\mathsf{B})$ is a coin-flipping protocol.*

It is common to also consider protocols with weaker correctness guarantee than the one we defined above, where with some small probability the output of the protocol in neither 0 nor 1. All the results we present in this paper can be easily generalized to handle such relaxations.

**Remark 4** ((partially) fair coin flipping)**.** *There are settings in which honest parties are required to always output a bit $c \in \{0,1\}$, even if the other party arbitrarily deviates from the prescribed protocol (and specifically, upon premature abort by the other party). Constructing coin-flipping protocols in this setting is a much more challenging task. Specifically, it is known that constructing an $m$-round $\delta$-bias coin-flipping protocol for $\delta \in o(1/m)$ is unconditionally impossible. We mention that since any $\delta$-bias coin-flipping protocol in this setting (i.e., with partial fairness) is also a $\delta$-bias coin-flipping protocol in our setting, our results hold for such protocols as well.*

A weaker variant of coin-flipping protocols (that we do not consider in this paper) is that of weak coin-flipping protocols. Such protocols are useful in the case that parties have (a priory known) opposite preferences.

**Definition 5.** *A polynomial-time protocol $(\mathsf{A},\mathsf{B})$ is a weak $\delta$-bias coin-flipping protocol, if the following hold:*

1. $\Pr[\mathrm{Out}(\mathsf{A},\mathsf{B})(n) = 0] = \Pr[\mathrm{Out}(\mathsf{A},\mathsf{B})(n) = 1] = \frac{1}{2}$, *and*
2. *there exist bits $c_{\mathsf{A}} \neq c_{\mathsf{B}} \in \{0,1\}$ such that the following holds for any* PPT*'s $\mathcal{A}$ and $\mathcal{B}$, and large enough $n$:*
   $\Pr[\mathrm{Out}(\mathcal{A},\mathsf{B})(n) = c_{\mathsf{A}}], \Pr[\mathrm{Out}(\mathsf{A},\mathcal{B})(n) = c_{\mathsf{B}}] \leq \frac{1}{2} + \delta(n)$.

*In the case that $\delta(n) = \mathrm{neg}(n)$, we simply say that $(\mathsf{A},\mathsf{B})$ is a weak coin-flipping protocol.*

## 2.3 One-Way Functions and Distributional One-Way Functions

An efficiently computable function is one-way if it is hard to invert it on a random output.

**Definition 6** (one-way functions)**.** *A polynomially-computable function $f : \{0,1\}^n \mapsto \{0,1\}^{\ell(n)}$ is* one-way*, if the following holds for any* PPT *$A$.*

$$\Pr_{y \leftarrow f(U_n)}[A(y) \in f^{-1}(y)] = \operatorname{neg}(n)$$

A seemingly weaker requirement is being distributional one-way, meaning that it is hard to sample a random preimage of a random output.

**Definition 7** (distributional one-way functions)**.** *A polynomially-computable function $f : \{0,1\}^n \mapsto \{0,1\}^{\ell(n)}$ is* distributional one-way*, if there exists $p \in \operatorname{poly}$ such that the following holds for any* PPT *$A$.*

$$\mathrm{SD}\left((U_n, f(U_n)), ((A(f(U'_n)), f(U'_n))\right) \geq \frac{1}{p(n)}$$

Clearly, any one-way function is also a distributional one-way function. While the other implication is not necessarily always true, Impagliazzo and Luby [16] showed that the existence of distributional one-way functions imply that of (standard) one-way functions. In particular, [16] proved that if one-way functions do not exists, then any efficiently computable function has an inverter of the following form.

**Definition 8** ($\gamma$-inverter)**.** *Let $f : \mathcal{D} \to \mathcal{R}$ be a deterministic function. An algorithm* Inv *is called a $\gamma$-inverter of $f$ the following holds.*

$$\mathrm{SD}\left((U, f(U)), (\mathsf{Inv}(f(U')), f(U'))\right) \leq \gamma,$$

*where $U, U'$ are uniformly distributed in $\mathcal{D}$.*

We call a 0-inverter of $f$, an *ideal* inverter of $f$. Alternatively, an ideal inverter of $f$ is an algorithm that on $y \in \mathcal{R}$, returns a uniformly chosen element (preimage) in $f^{-1}(y)$.

**Lemma 9** ([16, Lemma 1])**.** *Assume that one-way functions do not exit, then for any polynomial computable function $f : \{0,1\}^n \mapsto \{0,1\}^{\ell(n)}$ and any $p \in \operatorname{poly}$, there exists a* PPT Inv *that is a $1/p(n)$-inverter of $f$, for infinitely many $n$'s.*

Note that nothing is guaranteed when invoking a good inverter for $f$ (i.e., $\gamma$-inverter for some small $\gamma$) on an arbitrary distribution $D$. Yet, the following lemma states that if $D$ is dominated by output distribution of $f$, then such good inverters are useful.

**Lemma 10.** *Let $f : \mathcal{D} \to \mathcal{R}$ be a deterministic function and let* Ideal *be an ideal inverter of $f$. Let* A *be an oracle-aided algorithm that makes at most $m$ oracle queries to* Ideal*, where* *all* $\mathsf{A}$*'s queries are in* $\mathcal{R}$*. For* $i \in [m]$*, let the random variable* $Q_i$ *describe the* $i$*'th query of* $\mathsf{A}$*, where* $Q_i$ *is set to* $\perp$ *if the* $i$*'th query is not asked, and define the measure* $M_i$ *as follows:*

$$M_i(y) = \begin{cases} \Pr[Q_i = y] & y \in \mathcal{R}, \\ 0 & \textit{otherwise.} \end{cases}$$

*The probability is taken over the randomness of the algorithm* $\mathsf{A}$ *and the randomness of the ideal inverter* $\mathsf{Ideal}$*. Let* $U$ *denote the uniform distribution over* $\mathcal{D}$ *and suppose that* $f(U)$ $\delta$*-dominates* $M_i$ *for all* $i \in [m]$ *(according to Definition 2), then the following holds for any* $\gamma$*-inverter* $\mathsf{Inv}$ *of* $f$*.*

$$\mathrm{SD}\left(\mathsf{A}^{\mathsf{Ideal}}, \mathsf{A}^{\mathsf{Inv}}\right) \leq \frac{\gamma \cdot (m+1)}{\delta}.$$

*Proof.* We prove the lemma in two steps. In the first step, we prove that for the case that $m = 1$ it holds that

$$\mathrm{SD}\left(\mathsf{A}^{\mathsf{Ideal}}, \mathsf{A}^{\mathsf{Inv}}\right) \leq \frac{\gamma}{\delta}.$$

Then, we prove the case where $m > 1$, by reducing it back to the case of $m = 1$ (by considering a slightly different algorithm).

**The case of $\boldsymbol{m = 1}$.** Since the $\mathsf{A}$ behaves exactly the same in both cases up to the point in which the query is asked to the inverter, and since upon the same query/answer pair, the algorithm still behaves exactly the same in both cases, we may bound our discussion to the statistical distance between the distributions on query/answer pairs according to each random process. We denote by $Q$ the random variable describing the query (note, that this random variable is identically distributed in both scenarios). We denote by $X_{\mathsf{Ideal}}$ the random variable describing the answer given in an execution with the ideal inverter, and denote by $X_{\mathsf{Inv}}$ the random variable describing the answer given in an execution with the $\gamma$-inverter $\mathsf{Inv}$. Since in case $Q = \perp$ we are guaranteed that $X_{\mathsf{Ideal}} = X_{\mathsf{Inv}} = \perp$, we have that

$$\begin{aligned}
\mathrm{SD}\left(\mathsf{A}^{\mathsf{Ideal}}, \mathsf{A}^{\mathsf{Inv}}\right) &\leq \frac{1}{2} \cdot \sum_{q \in \mathcal{R}, a \in \{0,1\}^*} \left| \Pr_{\mathsf{Ideal}}[Q = q \wedge \mathsf{Ideal}(q) = a] - \Pr_{\mathsf{Inv}}[Q = q \wedge \mathsf{Inv}(q) = a] \right| \\
&= \frac{1}{2} \cdot \sum_{q \in \mathcal{R}, a \in \{0,1\}^*} \left| \Pr[Q = q] \cdot \Pr[\mathsf{Ideal}(q) = a] - \Pr[Q = q] \cdot \Pr[\mathsf{Inv}(q) = a] \right| \\
&= \frac{1}{2} \cdot \sum_{q \in \mathcal{R}, a \in \{0,1\}^*} \left| \Pr[Q = q] \cdot \left(\Pr[\mathsf{Ideal}(q) = a] - \Pr[\mathsf{Inv}(q) = a]\right) \right|.
\end{aligned}$$

Since $M_1$ is $\delta$-dominated by $f(U)$, we have that $\Pr[Q = q] = M_i(q) \leq \frac{\Pr[f(U)=q]}{\delta}$ for every

$q \in \mathcal{R}$. Hence,

$$\begin{aligned}
&\mathrm{SD}\left(\mathsf{A}^{\mathsf{Ideal}}, \mathsf{A}^{\mathsf{Inv}}\right) \\
&\leq \frac{1}{2} \cdot \sum_{q \in \mathcal{R}, a \in \{0,1\}^*} \left| \frac{\Pr[f(U) = q]}{\delta} \cdot (\Pr[\mathsf{Ideal}(q) = a] - \Pr[\mathsf{Inv}(q) = a]) \right| \\
&= \frac{1}{\delta} \cdot \frac{1}{2} \cdot \sum_{q \in \mathcal{R}, a \in \{0,1\}^*} |\Pr[f(U) = q] \cdot \Pr[\mathsf{Ideal}(q) = a] - \Pr[f(U) = q] \cdot \Pr[\mathsf{Inv}(q) = a])| \\
&= \frac{1}{\delta} \cdot \mathrm{SD}\left((U, f(U)), (\mathsf{Inv}(f(U')), f(U'))\right) \leq \frac{\gamma}{\delta}.
\end{aligned}$$

**The case of $m > 1$.** Define a sequence $\{H_i\}_{i=0}^m$ of hybrid random variables, by letting the $i$'th hybrid describe the output of the algorithm $\mathsf{A}$ in an execution where the first $i$ queries of $\mathsf{A}$ are answered by $\mathsf{Ideal}$ and the remaining $m - i$ queries are answered by $\mathsf{Inv}$. Specifically, we have that $H_m \equiv \mathsf{A}^{\mathsf{Ideal}}$ and that $H_0 \equiv \mathsf{A}^{\mathsf{Inv}}$. By the triangle inequality, we have that

$$\begin{aligned}
\mathrm{SD}\left(\mathsf{A}^{\mathsf{Ideal}}, \mathsf{A}^{\mathsf{Inv}}\right) &= \mathrm{SD}\left(H_0, H_m\right) \\
&\leq \sum_{i=1}^{m} \mathrm{SD}\left(H_{i-1}, H_i\right)
\end{aligned}$$

It, therefore, suffices to prove that for every $i$ it holds that $\mathrm{SD}\left(H_{i-1}, H_i\right) \leq \frac{\gamma}{\delta}$. Observe that except for the $i$'th query, both random variables describe the same random process, that can be viewed as an algorithm that asks only a single oracle query. Furthermore, since the first $i-1$ queries were asked to the ideal inverter, it holds that the $i$'th query is indeed distributed as $Q_i$. Thus, the lemma follows by applying the proof for the case where $m = 1$. □

# 3 The Attack

Let $(\mathsf{A}, \mathsf{B})$ be a coin-tossing protocol. In the following we define adversarial strategies for both $\mathsf{A}$ and $\mathsf{B}$ to bias the output of the protocol towards 1. The strategies for biasing the output towards 0 are defined analogously.

## 3.1 Notation

We associate the following random variables with an (honest) execution of $(\mathsf{A}, \mathsf{B})$. Throughout, we let $n$ be the security parameter of the protocol and omit it whenever its value is clear from the context. We assume for simplicity that the protocol's messages are single bits, and naturally view a valid execution of the protocol as a path the binary tree $\mathcal{T} = \mathcal{T}_n$, whose nodes are associated with all possible (valid) transcripts. The root of $\mathcal{T}$, corresponding to the empty transcript, is denoted by the empty string $\lambda$, and the children of a node $\alpha$ (if exist) are denoted by $\alpha \circ 0$ and $\alpha \circ 1$ ('$\circ$' stands for string concatenation), corresponding to the two (possibly partial) executions with these transcripts. A node with no descendants is called a *leaf*, where we assume for simplicity that a non-leaf node has exactly two descendants. Given a node $\alpha$, we let $|\alpha|$ denote its depth, and for $i \in [|\alpha|]$ let $\alpha_i$ denote the prefix of length $i$ of $\alpha$, which describes the $i$'th node on the path from $\lambda$ to $\alpha$ (e.g., $\alpha_0 = \lambda$).

We call a transcript $\alpha$ an $\mathsf{A}$ node [resp., $\mathsf{B}$ node], if this is $\mathsf{A}$'s [resp., $\mathsf{B}$'s] turn to send the next message, where without loss of generality the root $\lambda$ is an $\mathsf{A}$ node. We also assume that the parties always exchange $m = m(n)$ messages, and that each party uses $t = t(n)$ random coins, denoted $r_\mathsf{A}$ and $r_\mathsf{B}$ respectively. Given a pair of random coins $(r_\mathsf{A}, r_\mathsf{B})$, we let $\text{Leaf}(r_\mathsf{A}, r_\mathsf{B}) = (\mathsf{A}(r_\mathsf{A}), \mathsf{B}(r_\mathsf{B}))$ (i.e., the leaf transcript induced by the execution of $(\mathsf{A}(r_\mathsf{A}), \mathsf{B}(r_\mathsf{B}))$).

For $\alpha \in \mathcal{T}$, let $\mathsf{Uni}(\alpha)$ denote a random sample of $(r_\mathsf{A}, r_\mathsf{B})$, conditioned on $\text{Leaf}(r_\mathsf{A}, r_\mathsf{B})_{|\alpha|} = \alpha$. Given random coins $r_\mathsf{A} \in \{0,1\}^t$, we let $\mathsf{A}(r_\mathsf{A}; \alpha)$ be the next message sent by $\mathsf{A}$ with random coins $r_\mathsf{A}$ after seeing the transcript $\alpha$, and define the random variable $\mathsf{A}(\alpha)$ as $\mathsf{A}(R_\mathsf{A}; \alpha)$, where $(R_\mathsf{A}, *) \leftarrow \mathsf{Uni}(\alpha)$. [$\mathsf{B}(r_\mathsf{B}; \alpha)$ and $\mathsf{B}(\alpha)$ are defined analogously.] Finally, we assume without loss of generality that the transcript of an (honest) execution of the protocol always defines an output, 0 or 1 (consistent for both parties). For a leaf $\alpha$, we let $V_\alpha$ be the output of the protocol determined by this leaf, where if $\alpha$ is an internal node, we define $V_\alpha$ as

$$V_\alpha = \mathop{\mathrm{E}}_{(r_\mathsf{A}, r_\mathsf{B}) \leftarrow \mathsf{Uni}(\alpha)} [V_{\text{Leaf}(r_\mathsf{A}, r_\mathsf{B})}] \tag{2}$$

Namely, $V_\alpha$ is the probability that $(\mathsf{A}, \mathsf{B})$ outputs 1, conditioned that $\alpha$ is the current transcript.

Similarly, we associate the following random variables with an execution of $(\mathcal{A}, \mathsf{B})$, where $\mathcal{A}$ is a cheating strategy for $\mathsf{A}$: we denote the random coins used by $\mathcal{A}$ by $r_\mathcal{A}$, and for $\alpha \in \mathcal{T}$ let $\mathsf{Uni}^\mathcal{A}(\alpha)$ denote a random sample of $(r_\mathcal{A}, r_\mathsf{B})$, conditioned on $(\mathcal{A}(r_\mathcal{A}), \mathsf{B}(r_B))_{|\alpha|} = \alpha$. Given random coins $r_\mathcal{A} \in \{0,1\}^*$, we let $\mathcal{A}(r_\mathcal{A}; \alpha)$ be the next message sent by $\mathcal{A}$ with random coins $r_\mathcal{A}$ after seeing the transcript $\alpha$, and define the random variable $\mathcal{A}(\alpha)$ as $\mathcal{A}(R_\mathcal{A}; \alpha)$, where $(R_\mathcal{A}, *) \leftarrow \mathsf{Uni}^\mathcal{A}(\alpha)$. [$\mathcal{B}(r_\mathcal{B}; \alpha)$ and $\mathcal{B}(\alpha)$ are defined analogously.] Finally, we define $V_\alpha^\mathcal{A}$ as

$$V_\alpha^\mathcal{A} = \mathop{\mathrm{E}}_{(r_\mathcal{A}, r_\mathsf{B}) \leftarrow \mathsf{Uni}^\mathcal{A}(\alpha)} [V_{(\mathcal{A}(r_\mathcal{A}), \mathsf{B}(r_B))}], \tag{3}$$

where we set $V_{(\mathcal{A}(r_\mathcal{A}), \mathsf{B}(r_B))} = 0$, if $(\mathcal{A}(r_\mathcal{A}), \mathsf{B}(r_B))$ aborts. Namely, $V_\alpha^\mathcal{A}$ is a lower bound on the probability that $(\mathcal{A}, \mathsf{B})$ outputs 1, conditioned that $\alpha$ is the current transcript. [$V_\alpha^\mathcal{B}$ is defined analogously.]

## 3.2 The Adversary $\mathcal{A}$

We now present an adversarial strategy $\mathcal{A}$ for $\mathsf{A}$, designed to bias the outcome of the protocol towards 1 (the adversarial strategy $\mathcal{B}$ for $\mathsf{B}$ is defined analogously). In each round $\mathcal{A}$ uses a "sampling oracle" $\mathsf{Samp}$ to sample a value for the coins of $\mathsf{A}$, and then acts as the (honest) $\mathsf{A}$ would, given these coins and the current transcript. Roughly speaking, the objective of $\mathsf{Samp}$ is to return a random pair of coins $(r_\mathsf{A}, r_\mathsf{B})$ consistent with $\alpha$ (i.e., $\text{Leaf}(r_\mathsf{A}, r_\mathsf{B})_{|\alpha|} = \alpha$), which leads to a 1-node (i.e., $V_{\mathrm{Leaf}(r_{\mathsf{A}},r_{\mathsf{B}})} = 1$). In the following we analyze the success probability of $\mathcal{A}$ when using different implementations for $\mathsf{Samp}$. Specifically, in Section 4 we consider an "ideal sampler" (which is not necessarily efficient). Then, in Section 5, we consider a more realistic implementation of the sampler (specifically, using the inverter that will stem from the assumption that one-way function do not exist). Before describing and analyzing each of these samplers, we first give the formal description of $\mathcal{A}$.

**Algorithm 11** (Adversary $\mathcal{A}$)**.**

*Input: Security parameter $n$.*

*Oracle:* $\mathsf{Samp}$.

*Operation:* *Let $\alpha$ be the current transcript.*

```
1. Halt if $\alpha$ is a leaf node.
2. Let $(r_{\mathsf{A}}, *) \leftarrow \mathsf{Samp}(\alpha)$. Abort if $r_{\mathsf{A}} = \perp$.
3. Send $\mathsf{A}(r_{\mathsf{A}}; \alpha)$ to $\mathsf{B}$.
```

Given an instantiation of $\mathsf{Samp}$, we view $\mathcal{A}^{\mathsf{Samp}}$ as a random algorithm whose random coin are those used by $\mathsf{Samp}$ (independent coins for each call).

# 4 Using the Ideal Sampler

Our "ideal sampler" $\mathsf{Ideal}$ is defined as follows: on input $\alpha \in \mathcal{T}$, $\mathsf{Ideal}$ returns a random sample $(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)$, conditioned on $V_{\mathrm{Leaf}(r_{\mathsf{A}},r_{\mathsf{B}})} = 1$. Where $\mathsf{Ideal}$ returns $\perp$, in case $V_\alpha = 0$. The following lemma asserts that at least one of the parties has a good cheating strategy given oracle access to this sampler.

**Lemma 12.** *For any $n \in \mathbb{N}$ and any transcript $\alpha \in \mathcal{T}_n$, it holds that*

$$V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \cdot V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\alpha.$$

Assuming that $V_{\lambda_n} = 1/2$, it holds that either $V_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \geq 1/\sqrt{2}$ or $V_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} \geq 1/\sqrt{2}$. Namely, either $\mathcal{A}^{\mathsf{Ideal}}$ or $\mathcal{B}^{\mathsf{Ideal}}$ can bias the output of the protocol by $\frac{1}{\sqrt{2}} - \frac{1}{2}$.

*Proof of Lemma 12.* In following we prove the lemma using induction up the protocol tree. The proof is immediate for a leaf node and for an internal node $\alpha$ with $V_\alpha = 0$ (i.e., when $\mathsf{Ideal}(\alpha) = \perp$). In the following $\alpha$ is a fixed internal node (with $V_\alpha > 0$). Thus, for the sake of simplicity of notation, we let $\mathcal{A} = \mathcal{A}^{\mathsf{Ideal}}$, $V = V_\alpha$, and $V^{\mathcal{A}} = V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$. Similarly, for $j \in \{0,1\}$ we let $V_j = V_{\alpha j}$ and $V_j^{\mathcal{A}} = V_{\alpha j}^{\mathcal{A}^{\mathsf{Ideal}}}$. [$V^{\mathcal{B}}$ and $V_j^{\mathcal{B}}$ are defined analogously.] We need to prove that $V^{\mathcal{A}} \cdot V^{\mathcal{B}} \geq V$, assuming that $V_j^{\mathcal{A}} \cdot V_j^{\mathcal{B}} \geq V_j$ for both $j \in \{0,1\}$. We assume that $\alpha$ is an $\mathsf{A}$ node (the other case is analogous). Let $\beta = \Pr[\mathsf{A}(\alpha) = 1]$ (i.e., the probability that the next message of the honest $\mathsf{A}$ is 1). Note that $V = \beta \cdot V_1 + (1-\beta) \cdot V_0$, and that

$$\begin{aligned}
&\Pr[\mathcal{A}(\alpha) = 1] \\
&\quad= \Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)}[\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_{|\alpha|+1} = \alpha \circ 1 \mid V_{\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})} = 1] \\
&\quad= \frac{\Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)}[\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_{|\alpha|+1} = \alpha \circ 1 \wedge V_{\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})} = 1]}{\Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)}[V_{\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})} = 1]} \\
&\quad= \frac{\beta \cdot V_1}{V},
\end{aligned}$$

where the last equation is by a simple chain rule, i.e., since

$$\begin{aligned}
\beta &= \Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)}[\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_{|\alpha|+1} = \alpha \circ 1], \text{ and} \\
V_1 &= \Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)}[V_{\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})} = 1 \mid \mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_{|\alpha|+1} = \alpha \circ 1].
\end{aligned}$$

Thus,

- $V^{\mathcal{A}} = \Pr[\mathcal{A}(\alpha) = 1] \cdot V_1^{\mathcal{A}} + \Pr[\mathcal{A}(\alpha) = 0] \cdot V_0^{\mathcal{A}} = \frac{\beta \cdot V_1}{V} \cdot V_1^{\mathcal{A}} + \frac{(1-\beta) \cdot V_0}{V} \cdot V_0^{\mathcal{A}}$, and
- $V^{\mathcal{B}} = \Pr[\mathsf{A}(\alpha) = 1] \cdot V_1^{\mathcal{B}} + \Pr[\mathsf{A}(\alpha) = 0] \cdot V_0^{\mathcal{B}} = \beta \cdot V_1^{\mathcal{B}} + (1-\beta) \cdot V_0^{\mathcal{B}}$.

Using the induction hypothesis, we get that

$$\begin{aligned}
V^{\mathcal{A}} \cdot V^{\mathcal{B}} &= \left( \frac{\beta \cdot V_1}{V} \cdot V_1^{\mathcal{A}} + \frac{(1-\beta) \cdot V_0}{V} \cdot V_0^{\mathcal{A}} \right) \cdot \left( \beta \cdot V_1^{\mathcal{B}} + (1-\beta) \cdot V_0^{\mathcal{B}} \right) \\
&= \frac{\beta^2 \cdot V_1}{V} \cdot V_1^{\mathcal{A}} \cdot V_1^{\mathcal{B}} + \frac{(1-\beta)^2 \cdot V_0}{V} \cdot V_0^{\mathcal{A}} \cdot V_0^{\mathcal{B}} + \\
&\quad \frac{\beta(1-\beta)}{V} \cdot (V_1 \cdot V_1^{\mathcal{A}} \cdot V_0^{\mathcal{B}} + V_0 \cdot V_0^{\mathcal{A}} \cdot V_1^{\mathcal{B}}) \\
&\geq \frac{\beta^2 \cdot V_1^2}{V} + \frac{(1-\beta)^2 \cdot V_0^2}{V} + \frac{\beta \cdot (1-\beta) \cdot V_1 \cdot V_0}{V} \cdot \left( \frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}} + \frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}} \right).
\end{aligned}$$

Where since $V^2 = (\beta \cdot V_1 + (1-\beta) \cdot V_0)^2 = \beta^2 \cdot V_1^2 + (1-\beta)^2 \cdot V_0^2 + 2 \cdot \beta \cdot (1-\beta) \cdot V_1 \cdot V_0$, it follows that

$$\begin{aligned}
V^{\mathcal{A}} \cdot V^{\mathcal{B}} &\geq \frac{\beta^2 \cdot V_1^2}{V} + \frac{(1-\beta)^2 \cdot V_0^2}{V} + \frac{\beta \cdot (1-\beta) \cdot V_1 \cdot V_0}{V} \cdot \left( \frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}} + \frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}} \right) \\
&= V + \frac{\beta \cdot (1-\beta) \cdot V_1 \cdot V_0}{V} \cdot \left( \frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}} + \frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}} - 2 \right)
\end{aligned}$$

Thus, to prove that $V^{\mathcal{A}} \cdot V^{\mathcal{B}} \geq V$ it suffices to prove that $\frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}} + \frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}} \geq 2$. We will show that $\left( \frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}} + \frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}} - 2 \right) \cdot V_1^{\mathcal{A}} \cdot V_0^{\mathcal{A}} \geq 0$, which also suffices, since $V_1^{\mathcal{A}}$ and $V_0^{\mathcal{A}}$ are both positive values.

Indeed,

$$\begin{aligned}\left(\frac{V_1^{\mathcal{A}}}{V_0^{\mathcal{A}}}+\frac{V_0^{\mathcal{A}}}{V_1^{\mathcal{A}}}-2\right)\cdot V_1^{\mathcal{A}}\cdot V_0^{\mathcal{A}} &= (V_1^{\mathcal{A}})^2+(V_0^{\mathcal{A}})^2-2V_1^{\mathcal{A}}\cdot V_0^{\mathcal{A}}\\ &= (V_1^{\mathcal{A}}-V_0^{\mathcal{A}})^2\\ &\geq 0\end{aligned}$$

□

# 5 Moving to an Efficient Sampler

Our goal in this section is to use the above analysis of the success probability of our adversaries when given access to the *ideal* sampler, for analyzing their success probability when given access to an *efficient* sampler. The accuracy of such an inverter will be parameterized by a function $1/p$ for some $p \in \text{poly}$. In the following we fix such $p$.

Assuming that one-way functions do not exists, our efficient sampler is defined as follows: let $f : \{0,1\}^{t(n)} \times \{0,1\}^{t(n)} \times \{0,\ldots,m(n)\}$ be defined as

$$f(r_{\mathsf{A}}, r_{\mathsf{B}}, i) = \text{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_i, V_{\text{Leaf}(r_{\mathsf{A}},r_{\mathsf{B}})}$$

Namely, $f(r_{\mathsf{A}}, r_{\mathsf{B}}, i)$ outputs the $i$'th node in the execution of $(\mathsf{A}(r_{\mathsf{A}}), \mathsf{B}(r_{\mathsf{B}}))$ and the outcome coin induced by the leaf transcript of this execution. Given a node $\alpha \in \mathcal{T}_n$, the sampler $\mathsf{Real}_p$ returns $\mathsf{Inv}_f(\alpha, 1)$, where $\mathsf{Inv}_f$ is the distributional inverter for $f$ guaranteed by Lemma 9 with respect to accuracy parameter $1/p$.[3]

Notice that while Lemma 9 tells us that $\mathsf{Inv}_f$ samples well over a random output of $f$, the distribution induced by the calls of $\mathcal{A}^{\mathsf{Real}_p}$ might be very different from this distribution. While we cannot bound the difference between these two distributions, we prove that there exists a high-probability event conditioned upon these distributions are close enough. Loosely speaking, we first show that Lemma 12 still (almost) holds even if both $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ fail on their "non-typical" queries to $\mathsf{Ideal}$ – the calls that happen with provability very different for the one induce by $f$. Since $\mathsf{Real}_p$ does similarly to $\mathsf{Ideal}$ on the typical queries, it follows that $V_\lambda^{\mathcal{A}^{\mathsf{Real}_p}} \cdot V_\lambda^{\mathcal{B}^{\mathsf{Real}_p}}$ is almost as large as $V_\lambda$, and therefore, either $\mathcal{A}^{\mathsf{Real}_p}$ or $\mathcal{B}^{\mathsf{Real}_p}$ can significantly bias the outcome of the protocol.

In Lemma 13, stated below, we formally capture the above intuition regarding $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ (with access to the ideal sampler). We denote by $w(\alpha)$ the probability that the node $\alpha$ is visited in a random execution of $(\mathsf{A}, \mathsf{B})$ and by $w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)$ the probability of this visit in a random execution of $(\mathcal{A}^{\mathsf{Ideal}}, \mathsf{B})$. [$w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)$ is defined analogously.] Recall that we omit $n$ from the notation whenever its value is clear from the context. Specifically, we let $\lambda$ denote the root of $\mathcal{T}_n$, and $V_\lambda = \text{E}[\text{Out}(\mathsf{A}, \mathsf{B})(1^n)]$.

[3] We assume for simplicity that the security parameter of the protocol is determined by its (even partial) transcript, and therefore, the domain of $f$ in the calls to $\mathsf{Inv}_f$ is well defined.

**Lemma 13.** *Let* $(\mathsf{A},\mathsf{B})$ *be a coin-tossing protocol as above. For any* $q \in \operatorname{poly}$ *and for any* $n \in \mathbb{N}$*, there exists a set* $\mathcal{E} \subseteq \{\alpha \in \mathcal{T}_n \colon V_\alpha > 0\}$ *such that the following holds:*

1. *For any* $\alpha \in \mathcal{E}$*, it holds that* $\max\{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha), w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)\} \in O(q(n)^5 \cdot w(\alpha) \cdot V_\alpha)$*, and*
2. $V_\lambda^{{\mathcal{A}_\mathcal{E}}^{\mathsf{Ideal}}} \cdot V_\lambda^{{\mathcal{B}_\mathcal{E}}^{\mathsf{Ideal}}} \geq V_\lambda - \frac{1}{q(n)}$*,*

   *where* $\mathcal{A}_\mathcal{E}$ *acts as* $\mathcal{A}$ *does, but aborts if a node outside of* $\mathcal{E}$ *is reached. [*$\mathcal{B}_\mathcal{E}$ *is defined analogously.]*

Proving Lemma 13 is the main contribution of this section, but first let us use it for proving Theorem 14.

**Theorem 14** (restating Theorem 1)**.** *Let* $(\mathsf{A},\mathsf{B})$ *be a coin-tossing protocol with* $V_\lambda = \mathrm{E}[\mathrm{Out}(\mathsf{A},\mathsf{B})(1^n)]$*. Assuming that one-way functions do not exist, then for any* $g \in \operatorname{poly}$ *there exists a pair of efficient (cheating) strategies* $\mathcal{A}$ *and* $\mathcal{B}$ *such that the following holds: for infinitely many* $n$*'s, for each* $j \in \{0,1\}$ *either* $\Pr[(\mathcal{A}(j),\mathsf{B})(1^n) = j]$ *or* $\Pr[(\mathcal{B}(j),\mathsf{B})(1^n) = j]$ *is greater than* $\sqrt{V_n^j} - \frac{1}{g(n)}$*, where* $V_n^1 = V_\lambda$ *and* $V_n^0 = 1 - V_\lambda$*.*

In particular, for the case of $V_\lambda = \frac{1}{2}$, one party can "bias the outcome" of $(\mathsf{A},\mathsf{B})$ by almost $\frac{1}{\sqrt{2}} - \frac{1}{2}$.

*Proof.* We focus on $j = 1$ where the proof for $j = 0$ follows analogously. We prove the theorem by considering the success probabilities of the adversaries $\mathcal{A}^{\mathsf{Real}_p}$ and $\mathcal{B}^{\mathsf{Real}_p}$ (with access to an efficient inverter $\mathsf{Inv}_f$) on the set $\mathcal{E} \subseteq \mathcal{T}_n$ guaranteed by Lemma 13. Namely, the success probabilities of ${\mathcal{A}_\mathcal{E}}^{\mathsf{Real}_p}$ and ${\mathcal{B}_\mathcal{E}}^{\mathsf{Real}_p}$. We show that if $\mathsf{Inv}_f$ is "good enough", then they will do almost as well as ${\mathcal{A}_\mathcal{E}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_\mathcal{E}}^{\mathsf{Ideal}}$ would. Towards this end, we show that the distribution induced by $f$ on a random input, $(1/\operatorname{poly})$-dominates (according to Definition 2) both query distributions induced by ${\mathcal{A}_\mathcal{E}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_\mathcal{E}}^{\mathsf{Ideal}}$. Thus, we can apply Lemma 10 to show that each adversary behaves almost identically when given access to $\mathsf{Ideal}$ as when given access to $\mathsf{Inv}_f$. Finally, we remark that, while ${\mathcal{A}_\mathcal{E}}^{\mathsf{Real}_p}$ and ${\mathcal{B}_\mathcal{E}}^{\mathsf{Real}_p}$ may not be efficient (since they need to abort on $\alpha \notin \mathcal{E}$), they serve as a mental experiment and provide lower bounds on the success probabilities of $\mathcal{A}^{\mathsf{Real}_p}$ and $\mathcal{B}^{\mathsf{Real}_p}$, respectively. We next give the formal argument.

Let $g'(n) := \frac{g(n)}{\sqrt{V^1}}$, where we assume without loss of generality that $g(n) \geq \frac{1}{\sqrt{V^1}}$ (otherwise, the statement is trivial). Let $D_f(y)$ be the probability that a random output of $f$ equals $y$. Note that the following holds for any $\alpha \in \mathcal{T}_n$:

$$\begin{aligned} D_f(\alpha,1) &:= \Pr[f(U_{2t(n)}, I_n) = (\alpha,1)] \\ &= \Pr[I_n = |\alpha|] \cdot \Pr[\mathrm{Leaf}(U_{2t(n)})_{|\alpha|} = \alpha \wedge V_{\mathrm{Leaf}(U_{2t(n)})} = 1] \\ &= \frac{1}{m(n)+1} \cdot w(\alpha) \cdot V_\alpha \end{aligned}$$

where $I_n$ is uniformly distributed over $\{0,\ldots,m(n)\}$. Let $\mathcal{E} \subseteq \mathcal{T}_n$ be the set guaranteed by Lemma 13 with respect to $q(n) = 2 \cdot g'(n)$. It follows that

$$\max\{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha), w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)\} \in O(q(n)^5 \cdot m(n) \cdot D_f(\alpha,1)) \tag{4}$$

for any $\alpha \in \mathcal{E}$. In other words, the distributions induced by the queries of ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}$ on the range of $f$ are $\delta$-dominated by the distribution of a random output of $f$, for $\delta = 1/O(q(n)^5 \cdot m(n))$.

Fix $n \in \mathbb{N}$ such that the inverter $\mathsf{Inv}_f$ (guaranteed by Lemma 9) is a $1/p(n)$-inverter for $f$, and let $\mathsf{Real}_p$ be the sampler described above (i.e., $\mathsf{Real}_p(\alpha)$ returns $\mathsf{Inv}_f(\alpha, 1)$). For $\mathsf{Samp} \in \{\mathsf{Ideal}, \mathsf{Real}_p\}$, let $E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Samp}}}$ be the algorithm that emulates a random execution of $({\mathcal{A}_{\mathcal{E}}}^{\mathsf{Samp}}, \mathsf{B})$ and outputs the outcome of this execution, where $\mathcal{A}_{\mathcal{E}}$ is as in Lemma 13 [$E^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Samp}}}$ is defined analogously]. For $i \in \{0, \ldots, m(n)\}$, let $Q_i$ be the value of the $i$'th $\mathsf{Ideal}$-query made in the execution of $E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}}$ (set to $\perp$ if no such call was made). Equation (4) yields that for $\Pr[Q_i = (\alpha, 1)] \in O(q(n)^5 \cdot m(n) \cdot D_f(\alpha, 1))$ any $i \in [m(n)]$ and for any $\alpha \in \mathcal{E}$. Thus, Lemma 10 yields that

$$\mathrm{SD}(E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}}, E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}}) \in \frac{O\left(q(n)^5 \cdot m(n)^2\right)}{p(n)} < 1/8g'(n),$$

for the proper choice of $p$. Therefore,

$$V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}} = \Pr[E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}} = 1] \geq \Pr[E^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} = 1] - 1/4g'(n) = V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} - 1/4g'(n). \tag{5}$$

Doing the analogous calculation for $V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Real}_p}}$ and using Lemma 13, it follows that

$$\begin{aligned} V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}} \cdot V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Real}_p}} &\geq (V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} - 1/4g'(n)) \cdot (V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}} - 1/4g'(n)) \\ &= V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} \cdot V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}} - \frac{V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} + V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}}}{4g'(n)} + \frac{1}{2g'(n)^2} \\ &\geq V^1 - \frac{1}{q(n)} - \frac{1}{2g'(n)} = V^1 - \frac{1}{g'(n)}. \end{aligned} \tag{6}$$

Since $V_\lambda^{\mathcal{A}^{\mathsf{Real}_p}} \geq V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}}$ and $V_\lambda^{\mathcal{B}^{\mathsf{Real}_p}} \geq V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Real}_p}}$ (on the nodes in $\mathcal{E}$ the strategies ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}$ and $\mathcal{A}^{\mathsf{Real}_p}$ act identically, and ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Real}_p}$ fails on the other nodes), it follows that

$$V_\lambda^{\mathcal{A}^{\mathsf{Real}_p}} \cdot V_\lambda^{\mathcal{B}^{\mathsf{Real}_p}} \geq V^1 - \frac{1}{g'(n)}. \tag{7}$$

It follows that

$$V^1 - \frac{1}{g'(n)} = V^1 - \frac{\sqrt{V^1}}{g(n)} \geq V^1 - \frac{2 \cdot \sqrt{V^1}}{g(n)} + \frac{1}{g(n)^2} = \left(\sqrt{V^1} - \frac{1}{g(n)}\right)^2,$$

where the inequality holds since $g(n) \geq \frac{1}{\sqrt{V^1}}$. In particular, either $V_\lambda^{\mathcal{A}^{\mathsf{Real}_p}}$ or $V_\lambda^{\mathcal{B}^{\mathsf{Real}_p}}$ are larger than $\sqrt{V^1 - \frac{1}{g'(n)}} \geq \sqrt{V^1} - \frac{1}{g(n)}$, which completes the proof of the theorem. □

## 5.1 Proving Lemma 13

Towards proving Lemma 13 we identify the nodes (queries) in $\mathcal{T} = \mathcal{T}_n$ that are potentially "non typical" (i.e., either $V_\alpha$ is small or $\max\{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha), w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)\}$ is large), and prove that by modifying $\mathcal{A}^{\mathsf{Ideal}}$ or $\mathcal{B}^{\mathsf{Ideal}}$ to totally fail on such nodes, we hardly change their overall success probability. Alternatively, if $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ abort whenever they reach a non-typical node (as do ${\mathcal{A}_\mathcal{E}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_\mathcal{E}}^{\mathsf{Ideal}}$), then they will only give away a $1/\operatorname{poly}$ fraction of their success probability. The proof then follows by taking $\mathcal{E}$ to be the set of "typical" nodes in $\mathcal{T}$.

We next give a slightly more detailed overview of the proof. For simplicity, in the discussion below, we (implicitly) assume that $V_\lambda$ is constant (in the formal proof, we deal with any value of $V_\lambda$). We need to show that the set $\mathcal{E}$ satisfies both of the requirements in Lemma 13. Proving that the first requirement is satisfied will come for free, simply by the way we define non-typical nodes. To show that $\mathcal{E}$ satisfies the second requirement (i.e., that $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ can indeed abort on nodes outside $\mathcal{E}$ without losing much), we partition the non-typical nodes into two sets. The first set, denoted $\mathsf{Small}$, contains those nodes for which $V_\alpha \in O(\frac{1}{q^2})$. The second set, denoted $\mathsf{UnBal}$, contains the nodes whose weights induced by $\mathcal{A}^{\mathsf{Ideal}}$ or $\mathcal{B}^{\mathsf{Ideal}}$ are $\Omega(q^2)$ times larger then their weight in an honest execution of the protocol. On a very intuitive level, handling the set $\mathsf{Small}$ is fairly easy: consider a mental experiment in which we (artificially) set a new "success probability" for such nodes, by setting $V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} = \sqrt{V_\alpha}$ for every $\alpha \in \mathsf{Small}$. Since $V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \cdot V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\alpha$, the proof of Lemma 12 will still go through with respect to the above experiment. Namely, it will still hold that $V_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \cdot V_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\lambda$. To then allow aborting on nodes in $\mathsf{Small}$, we observe that neither $\mathcal{A}^{\mathsf{Ideal}}$ nor $\mathcal{B}^{\mathsf{Ideal}}$ gains much on any node $\alpha \in \mathsf{Small}$ (at most $\sqrt{V_\alpha} \in O(1/q)$). Hence, even if $\mathsf{Small}$ is reached with high probability, it contributes an overall success probability of $O(1/q)$.

Handling the unbalanced nodes inside $\mathsf{UnBal}$, on the other hand, seems much more challenging. These nodes might have arbitrary expected values (i.e., $V_\alpha$) and are reached by one of the adversaries with high probability. As such, they may contribute significantly to the success probability of the cheating parties. Fortunately, by making a critical use of the query distribution induced by the ideal sampler, we are able to prove the following "compensation lemma": a node $\alpha$ whose weight with respect to $\mathcal{A}^{\mathsf{Ideal}}$ is $k$ times *larger* from its real weight (i.e., $w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = k \cdot w(\alpha)$), has weight with respect to $\mathcal{B}^{\mathsf{Ideal}}$ that is $k$ time *smaller* than its real weight. Hence, the set $\mathsf{UnBal}$ can be separated into two disjoint subsets $\mathsf{UnBal}_\mathcal{A}$ and $\mathsf{UnBal}_\mathcal{B}$, where $\mathsf{UnBal}_\mathcal{B}$ is almost never visited by $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathsf{UnBal}_\mathcal{A}$ is almost never visited by $\mathcal{B}^{\mathsf{Ideal}}$. Now, we handle each of these sets in a similar manner to the way we handled the nodes in $\mathsf{Small}$ (for simplicity we only consider here the set $\mathsf{UnBal}_\mathcal{A}$): consider the mental experiment in which for every $\alpha \in \mathsf{UnBal}_\mathcal{A}$ we modify the values of $V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$ and $V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}}$ such that $V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = 1/q$ and $V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} = q$ (this is only a mental experiment, so we do not care that these values might be larger than 1). Since $V_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \cdot V_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} = 1 \geq V_\alpha$, the proof of Lemma 12 still goes through with respect to this experiment as well. Furthermore, we can safely fail both cheating strategies on $\mathsf{UnBal}_\mathcal{A}$ without changing their overall success probability too much. Specifically, $\mathcal{A}^{\mathsf{Ideal}}$ will not suffer much because its success probability on these nodes is bounded by $\frac{1}{q}$ (i.e., it has gained at most $O(1 \cdot \frac{1}{q} = \frac{1}{q})$ from these nodes), and $\mathcal{B}^{\mathsf{Ideal}}$ will not suffer much since it almost never visits these nodes (i.e., it has gained $O(q \cdot \frac{1}{q^2} = \frac{1}{q})$ from these nodes).

We now work towards formalizing the above discussion. We assume that $V_\lambda \geq 1/q$, since otherwise the lemma follows trivially, and start with formally defining the different subsets of $\mathcal{T}$ we considered above. We define the relative weights of $\alpha \in \mathcal{T}$ as $W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = \frac{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)}$ and $W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) = \frac{w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)}$, let

$$\mathsf{UnBal}_{\mathcal{A}} := \{\alpha \in \mathcal{T} \colon W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) > 16 \cdot q^3\} \tag{8}$$

$$\mathsf{UnBal}_{\mathcal{B}} := \{\alpha \in \mathcal{T} \colon W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) > 16 \cdot q^3\}, \tag{9}$$

and let $\mathsf{UnBal} = \mathsf{UnBal}_{\mathcal{A}} \cup \mathsf{UnBal}_{\mathcal{B}}$. Finally, we let

$$\mathsf{Small} := \{\alpha \in \mathcal{T} \setminus \mathsf{UnBal} \colon V_\alpha < \frac{1}{16 \cdot q^2}\} \tag{10}$$

and let $\mathcal{E} = \mathcal{T} \setminus (\mathsf{Small} \cup \mathsf{UnBal})$. The following fact is immediate.

**Claim 15.** *For any $\alpha \in \mathcal{E}$ it holds that* $\max\{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha), w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)\} \in O(q^5 \cdot w(\alpha) \cdot V_\alpha)$.

*Proof.* For any $\alpha \in \mathcal{E}$, it holds that $\alpha \notin \mathsf{UnBal}_{\mathcal{A}}$. Hence,

$$w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot w(\alpha) \leq 16 \cdot q^3 \cdot w(\alpha) \leq 2^8 \cdot q^5 \cdot w(\alpha) \cdot V_\alpha,$$

where the last inequality follows since $\alpha \notin \mathsf{Small}$ (and therefore $16 \cdot q^2 \cdot V_\alpha \geq 1$). □

To prove that $\mathcal{E}$ satisfies the second property of Lemma 13, we present a pair of random variables $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$ and $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$, such that the following holds for $\lambda$ (the root of $\mathcal{T}$):

1. $Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\lambda$, and
2. $V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} - 1/2q$ and $V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - 1/2q$.

The variables $Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}}$ and $Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}}$ are defined below, but intuitively they measure the success probability of $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ respectively, in the mental experiment where their success probability on internal nodes outside $\mathcal{E}$ is changed according to the informal description above. The above immediately yields that $V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} \cdot V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}} \geq V_\lambda - \frac{1}{q}$, completing the proof of Lemma 13.

Since our goal is to bound (from below) the success probabilities of ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}$, it suffices to restrict the discussion to the nodes in $\mathcal{T}$ that have non-zero probability of being reached in executions with ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}$. This set of nodes defines a tree (which is defined below and denoted $\mathcal{T}'$) that can alternatively be defined as the set of all nodes in $\mathcal{T}$ that have no proper ancestor in $\mathsf{Small} \cup \mathsf{UnBal}$. We use the following random variables:

**Definition 16.** *For $\alpha \in \mathcal{T}' := \operatorname{Supp}((\mathcal{A}, \mathsf{B})(1^n)) \cap \operatorname{Supp}((\mathsf{A}, \mathcal{B})(1^n)) \subseteq \mathcal{T}$,*[4] *we define $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$ as follows [$Y_\alpha^{\mathcal{B}^{\mathsf{Ideal}}}$ is defined analogously]:*

[4] We assume without loss of generality that an honest party aborts if the other party does. Hence, $\mathcal{T}'$ is indeed contained in $\mathcal{T}$.

- *If $\alpha \in \mathcal{E}$:*
    1. *If $\alpha$ is a leaf, $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = V_\alpha$.*
    2. *Otherwise, $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = \Pr[\mathcal{A}^{\mathsf{Ideal}}(\alpha) = 1] \cdot Y_{\alpha\circ 1}^{\mathcal{A}^{\mathsf{Ideal}}} + \Pr[\mathcal{A}^{\mathsf{Ideal}}(\alpha) = 0] \cdot Y_{\alpha\circ 0}^{\mathcal{A}^{\mathsf{Ideal}}}$.*
- *If $\alpha \in \mathsf{UnBal}$:*
    1. *If $\alpha \in \mathsf{UnBal}_{\mathcal{A}}$, $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = \frac{1}{4q}$.*
    2. *Otherwise ($\alpha \in \mathsf{UnBal}_{\mathcal{B}}$), $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = 4q$.*
- *Otherwise ($\alpha \in \mathsf{Small}$), $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = \frac{1}{4q}$.*

We emphasize that the adversaries $\mathcal{A}^{\mathsf{Ideal}}$ and $\mathcal{B}^{\mathsf{Ideal}}$ remain exactly as before, and the random variables $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}$ and $Y_\alpha^{\mathcal{B}^{\mathsf{Ideal}}}$ only enable us to present a refined analysis of their success probabilities. The following fact easily follows from similar arguments to those used in the proof of Lemma 12.

**Claim 17.** *For any $\alpha \in \mathcal{T}'$, it holds that*

$$Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\alpha.$$

*Proof.* The proof is by induction up the protocol tree. For a node $\alpha \notin \mathcal{E}$, the lemma is trivially true since $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\alpha$. For any other node $\alpha$ (without loss of generality $\alpha$ is an $\mathsf{A}$ node), the proof follows from exactly the same argument as in Lemma 12. This is true since for the base cases nothing has changed, and for an internal node $\alpha$ it holds that

- $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = \Pr[\mathcal{A}^{\mathsf{Ideal}}(\alpha) = 1] \cdot Y_{\alpha\circ 1}^{\mathcal{A}^{\mathsf{Ideal}}} + \Pr[\mathcal{A}^{\mathsf{Ideal}}(\alpha) = 0] \cdot Y_{\alpha\circ 0}^{\mathcal{A}^{\mathsf{Ideal}}}$, and
- $Y_\alpha^{\mathcal{B}^{\mathsf{Ideal}}} = \Pr[\mathsf{A}(\alpha) = 1] \cdot Y_{\alpha\circ 1}^{\mathcal{B}^{\mathsf{Ideal}}} + \Pr[\mathsf{A}(\alpha) = 0] \cdot Y_{\alpha\circ 0}^{\mathcal{B}^{\mathsf{Ideal}}}$.

Hence, the proof of the induction step follows exactly as in the proof of Lemma 12, which uses no property of the children of $\alpha$ other than that they satisfy the induction hypothesis. □

To complete the proof of Lemma 13, we need to prove that the success probability of both ${\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}$ and ${\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}$ is not far from the above mental experiment. We prove the following lemma.

**Lemma 18.** *It holds that $V_\lambda^{{\mathcal{A}_{\mathcal{E}}}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} - 1/2q$ and $V_\lambda^{{\mathcal{B}_{\mathcal{E}}}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - 1/2q$.*

*Proof.* The main tool we are using for proving Lemma 18 is the following "compensation lemma".

**Lemma 19** (compensation lemma)**.** *Let the relative weights of $\alpha \in \mathcal{T}$ be as above (i.e., $W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = \frac{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)}$ and $W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) = \frac{w^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)}$). The following holds for every $\alpha \in \mathcal{T}$:*

$$W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) = \frac{V_\alpha}{V_\lambda}.$$

Namely, the lemma states that a node $\alpha$ whose weight with respect to $\mathcal{A}^{\mathsf{Ideal}}$ is $k$ times larger its typical weight (i.e., $w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) > k \cdot w(\alpha)$), has weight with respect to $\mathcal{B}^{\mathsf{Ideal}}$ that is (close to) $k$ times smaller then its typical weight. The proof of Lemma 19 is given later below. We first use it for completing the proof of Lemma 18.

In the following we focus on analyzing the value of $V_\lambda^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}}$ (the part of $V_\lambda^{\mathcal{B}_\mathcal{E}{}^{\mathsf{Ideal}}}$ is proved analogously). Let $\mathcal{F}$ be the set of leaves in $\mathcal{T}'$. That is, $\mathcal{F}$ contains nodes of two types: (i) a leaf $\alpha$ of the original tree $\mathcal{T}$ (such that, there is no ancestor $\alpha'$ of $\alpha$ in $\mathsf{Small} \cup \mathsf{UnBal}$), and (ii) a node $\alpha \in \mathsf{Small} \cup \mathsf{UnBal}$ (such that, there is no ancestor $\alpha' \neq \alpha$ of $\alpha$ in $\mathsf{Small} \cup \mathsf{UnBal}$). Furthermore, any execution $(\mathcal{A}^{\mathsf{Ideal}}, \mathsf{B})$ passes through a node in $\mathcal{F}$. It follows that

$$Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} = \sum_{\alpha \in \mathcal{F}} w^{\mathcal{A}}(\alpha) \cdot Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \tag{11}$$

$$V_\lambda^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}} = \sum_{\alpha \in \mathcal{F}} w^{\mathcal{A}}(\alpha) \cdot V_\alpha^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}} \tag{12}$$

Let

- $\mathcal{F}_1 = \mathcal{F} \cap \mathsf{UnBal}_\mathcal{B}$,
- $\mathcal{F}_2 = \mathcal{F} \cap (\mathsf{UnBal}_\mathcal{A} \cup \mathsf{Small})$, and
- $\mathcal{F}_3 = \mathcal{F} \setminus (\mathcal{F}_1 \cup \mathcal{F}_2) = \mathcal{F} \cap \mathcal{E}$.

Lemma 19 yields that $\mathsf{UnBal}_\mathcal{A}$ and $\mathsf{UnBal}_\mathcal{B}$ are disjoint. It follows that $\mathcal{F}_1, \mathcal{F}_2$, and $\mathcal{F}_3$ form a partition of $\mathcal{F}$, and Equation (11) yields that

$$\begin{aligned} Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} &= \sum_{\alpha \in \mathcal{F}_1} w^{\mathcal{A}}(\alpha) \cdot Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} + \sum_{\alpha \in \mathcal{F}_2} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} + \sum_{\alpha \in \mathcal{F}_3} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} \\ &\leq \sum_{\alpha \in \mathcal{F}_1} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot 4q + \sum_{\alpha \in \mathcal{F}_2} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot \frac{1}{4q} + \sum_{\alpha \in \mathcal{F}_3} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot V_\alpha^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}} \\ &\leq 4q \cdot \sum_{\alpha \in \mathcal{F}_1} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) + \frac{1}{4q} \cdot \sum_{\alpha \in \mathcal{F}_2} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) + V_\lambda^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}} \\ &\leq 4q \cdot \sum_{\alpha \in \mathcal{F}_1} w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) + \frac{1}{4q} + V_\lambda^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}}, \end{aligned} \tag{13}$$

where the first inequality follows from Definition 16, which yields that $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = 4q$ for any $\alpha \in \mathcal{F}_1$, that $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = 1/4q$ for any $\alpha \in \mathcal{F}_2$ and that $Y_\alpha^{\mathcal{A}^{\mathsf{Ideal}}} = V_\alpha^{\mathcal{A}_\mathcal{E}{}^{\mathsf{Ideal}}}$ for any $\alpha \in \mathcal{F}_3$. The second inequality follows from Equation (12).

We next consider the probability of visiting $\mathcal{F}_1$ in a random an execution of $(\mathcal{A}^{\mathsf{Ideal}}, \mathsf{B})$. The definition of $\mathsf{UnBal}_\mathcal{B}$ yields that $W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) > 16 \cdot q^3$ for any $\alpha \in \mathcal{F}_1$. Applying Lemma 19 yields that

$$\frac{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)} = W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) < \frac{1}{16 \cdot q^2} \cdot \frac{V_\alpha}{V_\lambda}, \tag{14}$$

for any $\alpha \in \mathcal{F}_1$. Since $V_\alpha \leq 1$ and $\frac{1}{V_\lambda} \leq q$, we have that $w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) < \frac{w(\alpha)}{16 \cdot q^2}$. Plugging this into Equation (13) yields that

$$\begin{aligned}
Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} &< 4q \cdot \sum_{\alpha \in \mathcal{F}_1} \frac{w(\alpha)}{16 \cdot q^2} \; + \; \frac{1}{4q} \; + \; V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \\
&= \frac{4q}{16 \cdot q^2} \cdot \sum_{\alpha \in \mathcal{F}_1} w(\alpha) \; + \; \frac{1}{4q} \; + \; V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \\
&\leq \frac{1}{4 \cdot q} \; + \; \frac{1}{4 \cdot q} \; + \; V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \\
&= \frac{1}{2 \cdot q} \; + \; V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}}
\end{aligned}$$

Hence, $V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} - 1/2q$, as desired. □

### 5.1.1 Putting it All Together.

We next summarize the arguments that conclude the proof of Lemma 13.

*Proof of Lemma 13.* Let $\mathcal{E}$ be defined as in the foregoing discussion. Claim 15 asserts that $\mathcal{E}$ satisfies the first requirement of Lemma 13. For the second requirement, Lemma 18 yields that $V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} - \frac{1}{2q}$ and $V_\lambda^{\mathcal{B}_\mathcal{E}^{\mathsf{Ideal}}} \geq Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - \frac{1}{2q}$. Hence, we have

$$\begin{aligned}
V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \cdot V_\lambda^{\mathcal{B}_\mathcal{E}^{\mathsf{Ideal}}} &\geq (Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} - \frac{1}{2q}) \cdot (Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - \frac{1}{2q}) \\
&= Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - \frac{Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} + Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}}}{2q} + \left(\frac{1}{2q}\right)^2 \\
&\geq Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} - \frac{2}{2q}.
\end{aligned}$$

Claim 17 asserts that $Y_\lambda^{\mathcal{A}^{\mathsf{Ideal}}} \cdot Y_\lambda^{\mathcal{B}^{\mathsf{Ideal}}} \geq V_\lambda$, and hence, the second requirement is also satisfied, i.e.,

$$V_\lambda^{\mathcal{A}_\mathcal{E}^{\mathsf{Ideal}}} \cdot V_\lambda^{\mathcal{B}_\mathcal{E}^{\mathsf{Ideal}}} \geq V_\lambda - \frac{1}{q}.$$

□

### 5.1.2 The Proof of the Compensation Lemma.

*Proof of Lemma 19.* For $\alpha \in \mathcal{T}$ and $c \in \{0,1\}$, let $\beta_\alpha(c)$ be the probability that the next message is $c$ given that the transcript so far was $\alpha$. I.e.,

$$\beta_\alpha(c) = \Pr_{(r_{\mathsf{A}}, r_{\mathsf{B}}) \leftarrow \mathsf{Uni}(\alpha)} [\mathrm{Leaf}(r_{\mathsf{A}}, r_{\mathsf{B}})_{|\alpha|+1} = \alpha \circ c] \qquad (15)$$

Recall that $w(\alpha)$ is the probability that $\alpha$ is a prefix of the full communication transcript in an honest execution of the protocol. Assume that $\alpha = c_1 c_2 \ldots c_\ell$, then $w(\alpha) = \beta_{\alpha_0}(c_1) \cdot \beta_{\alpha_1}(c_2) \cdot \ldots \cdot \beta_{\alpha_{\ell-1}}(c_\ell)$.

Consider now an execution of $(\mathcal{A}^{\mathsf{Ideal}}, \mathsf{B})$. For $c \in \{0,1\}$, let $\beta_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}(c)$ be the probability that the next message is $c$ given that the transcript so far was $\alpha$. I.e.,

$$\beta_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}(c) = \Pr[\mathcal{A}^{\mathsf{Ideal}}(\alpha) = c]. \tag{16}$$

Recall that $w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)$ is the probability that the node $\alpha$ is reached in an execution of $(\mathcal{A}^{\mathsf{Ideal}}, \mathsf{B})$. It follows that

$$w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = \beta_{\alpha_0}^{\mathcal{A}^{\mathsf{Ideal}}}(c_1) \cdot \beta_{\alpha_1}^{\mathcal{A}^{\mathsf{Ideal}}}(c_2) \cdots \beta_{\alpha_{\ell-1}}^{\mathcal{A}^{\mathsf{Ideal}}}(c_\ell),$$

Note that, if $\alpha$ is an $\mathsf{A}$ node, then $\beta_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}(c) = \frac{\beta_\alpha(c) \cdot V_{\alpha \circ c}}{V_\alpha}$, and otherwise $\beta_\alpha^{\mathcal{A}^{\mathsf{Ideal}}}(c) = \beta_\alpha(c)$. It follows that

$$\begin{aligned}
W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) &= \frac{w^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha)}{w(\alpha)} \\
&= \frac{1}{w(\alpha)} \cdot \beta_{\alpha_0}^{\mathcal{A}^{\mathsf{Ideal}}}(c_1) \beta_{\alpha_1}^{\mathcal{A}^{\mathsf{Ideal}}}(c_2) \cdot \ldots \cdot \beta_{\alpha_{\ell-2}}^{\mathcal{A}^{\mathsf{Ideal}}}(c_{\ell-1}) \cdot \beta_{\alpha_{\ell-1}}^{\mathcal{A}^{\mathsf{Ideal}}}(c_\ell) \\
&= \frac{1}{w(\alpha)} \cdot \frac{\beta_{\alpha_0}(c_1) \cdot V_{\alpha_1}}{V_{\alpha_0}} \beta_{\alpha_1}(c_2) \cdot \ldots \cdot \frac{\beta_{\alpha_{\ell-2}}(c_{\ell-1}) \cdot V_{\alpha_{\ell-1}}}{V_{\alpha_{\ell-2}}} \beta_{\alpha_{\ell-1}}(c_\ell) \\
&= \frac{\beta_{\alpha_0}(c_1) \beta_{\alpha_1}(c_2) \cdot \ldots \cdot \beta_{\alpha_{\ell-2}}(c_{\ell-1}) \beta_{\alpha_{\ell-1}}(c_\ell)}{w(\alpha)} \cdot \frac{V_{\alpha_1}}{V_{\alpha_0}} \cdot 1 \cdot \ldots \cdot \frac{V_{\alpha_{\ell-1}}}{V_{\alpha_{\ell-2}}} \cdot 1
\end{aligned}$$

Since $w(\alpha) = \beta_{\alpha_0}(c_1) \beta_{\alpha_1}(c_2) \cdot \ldots \cdot \beta_{\alpha_{\ell-2}}(c_{\ell-1}) \beta_{\alpha_{\ell-1}}(c_\ell)$, we obtain that

$$W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) = \prod_{i=1}^{\ell/2} \frac{V_{\alpha_{2i-1}}}{V_{\alpha_{2i-2}}}.$$

A similar argument shows that

$$W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) = \prod_{i=1}^{\ell/2} \frac{V_{\alpha_{2i}}}{V_{\alpha_{2i-1}}}.$$

Hence, we conclude that

$$W^{\mathcal{A}^{\mathsf{Ideal}}}(\alpha) \cdot W^{\mathcal{B}^{\mathsf{Ideal}}}(\alpha) = \prod_{i=1}^{\ell/2} \frac{V_{\alpha_{2i-1}}}{V_{\alpha_{2i-2}}} \cdot \prod_{i=1}^{\ell/2} \frac{V_{\alpha_{2i}}}{V_{\alpha_{2i-1}}} = \prod_{i=1}^{\ell} \frac{V_{\alpha_i}}{V_{\alpha_{i-1}}} = \frac{V_\alpha}{V_\lambda}$$

□

# 6 Discussion and Open Questions

The main open question is understanding the limits of efficient attacks in breaking coin-flipping protocols. Specifically (assuming one-way functions do not exist), does there exist, for any (correct) coin-flipping protocol, an efficient adversary that biases its output towards 0 *or* towards 1 by $\frac{1}{2} - 1/\operatorname{poly}$? or even by $\frac{\sqrt{2}-1}{2} + \Omega(1)$? In light of the reduction of Chailloux and Kerenidis [4] from $(\frac{\sqrt{2}-1}{2} + O(\varepsilon))$-bias *strong* coin-flipping to $\varepsilon$-bias *weak* coin-flipping, a positive answer (even to the weaker form of above question, i.e., $\frac{\sqrt{2}-1}{2} + \Omega(1)$ bias), would imply that the existence of constant-bias weak coin-flipping protocols implies the existence of one-way functions.

While our analysis only proves the existence of an adversary achieving $\frac{\sqrt{2}-1}{2} - o(1)$ bias (and thus has no direct implication to weak coin flipping), it shows that (assuming one-way functions do not exist) for any coin-flipping protocol there exists an efficient adversary that can bias its output *both* towards 0 and towards 1, by $\frac{\sqrt{2}-1}{2} - o(1)$. Hence, our attack accomplishes a harder task than the required one. Interestingly, $\frac{\sqrt{2}-1}{2}$ is the right bound for this more challenging task. That is, there exists a (correct) coin-flipping protocol for which no adversary (not even an unbounded one) can bias the output towards 1 by more than $\frac{\sqrt{2}-1}{2}$.[5]

## Acknowledgment

We thank Hemanta Maji for useful discussions.

## References

[1] B. Averbuch, M. Blum, B. Chor, S. Goldwasser, and S. Micali. How to implement Bracha's $O(\log n)$ Byzantine agreement algorithm, 1985. Unpublishe manuscript.

[2] A. Beimel, E. Omri, and I. Orlov. Protocols for multiparty coin toss with dishonest majority. In T. Rabin, editor, *Advances in Cryptology – CRYPTO 2010*, volume 6223 of *Lecture Notes in Computer Science*, pages 538–557. Springer, 2010.

[3] M. Blum. Coin flipping by telephone. In *Advances in Cryptology – CRYPTO '81*, pages 11–15, 1981.

[4] A. Chailloux and I. Kerenidis. Optimal quantum strong coin flipping. In *Proceedings of the 50th Annual Symposium on Foundations of Computer Science (FOCS)*, pages

[5] For instance, consider the protocol where $\mathsf{A}$ (playing first) sets the outcome of the protocol to 0 w.p. $1 - \frac{1}{\sqrt{2}}$ and defers the decision to $\mathsf{B}$ otherwise. The party $\mathsf{B}$, if plays, sets the outcome to 1 w.p. $\frac{1}{\sqrt{2}}$ and to 0 otherwise. It is clear that the protocol is correct (i.e., expected outcome for an honest execution is $\frac{1}{2}$) and that there exists no cheating strategy for neither $\mathsf{A}$ nor $\mathsf{B}$ that can make the expected outcome of the protocol to be larger than $\frac{1}{\sqrt{2}}$.

527–533, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3850-1. doi: http://dx.doi.org/10.1109/FOCS.2009.71. URL http://dx.doi.org/10.1109/FOCS.2009.71.

[5] K.-M. Chung and F.-H. Liu. Parallel repetition theorems for interactive arguments. In *Theory of Cryptography, Fourth Theory of Cryptography Conference, TCC 2010*, pages 19–36, 2010.

[6] R. Cleve. Limits on the security of coin flips when half the processors are faulty. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 364–369, 1986.

[7] R. Cleve and R. Impagliazzo. Martingales, collective coin flipping and discrete control processes. Manuscript, 1993.

[8] D. Dachman-Soled, Y. Lindell, M. Mahmoody, and T. Malkin. On the black-box complexity of optimally-fair coin tossing. In *Theory of Cryptography, Fourth Theory of Cryptography Conference, TCC 2011*, volume 6597, pages 450–467, 2011.

[9] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 25–32, 1989.

[10] O. Goldreich, S. Goldwasser, and S. Micali. On the cryptographic applications of random functions. In *Advances in Cryptology – CRYPTO '84*, pages 276–288, 1984.

[11] O. Goldreich, S. Goldwasser, and S. Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, 1986.

[12] I. Haitner. A parallel repetition theorem for any interactive argument. In *Proceedings of the 50th Annual Symposium on Foundations of Computer Science (FOCS)*, 2009.

[13] I. Haitner, M. Nguyen, S. J. Ong, O. Reingold, and S. Vadhan. Statistically hiding commitments and statistical zero-knowledge arguments from any one-way function. *SIAM Journal on Computing*, 39(3):1153–1218, 2009.

[14] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. Preliminary versions in *STOC'89* and *STOC'90*.

[15] J. Håstad, R. Pass, K. Pietrzak, and D. Wikström. An efficient parallel repetition theorem. In *Theory of Cryptography, Fourth Theory of Cryptography Conference, TCC 2010*, 2010.

[16] R. Impagliazzo and M. Luby. One-way functions are essential for complexity based cryptography. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 230–235, 1989.

[17] A. Y. Kitaev. Quantum coin-flipping. presentation at the 6th workshop on quantum information processing (qip 2003), 2003.

[18] H. K. Maji, M. Prabhakaran, and A. Sahai. On the Computational Complexity of Coin Flipping. In *Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 613–622, 2010.

[19] C. Mochon. Quantum weak coin flipping with arbitrarily small bias. arXiv:0711.4114, 2007.

[20] T. Moran, M. Naor, and G. Segev. An optimally fair coin toss. In *Theory of Cryptography, Fourth Theory of Cryptography Conference, TCC 2009*, pages 1–18, 2009.

[21] M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991. Preliminary version in *CRYPTO'89*.

[22] M. Naor and M. Yung. Universal one-way hash functions and their cryptographic applications. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 33–43. ACM Press, 1989.

[23] J. Rompel. One-way functions are necessary and sufficient for secure signatures. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 387–394, 1990.